Exploiting NOMA into Socially Enabled Computation Offloading

Yutong Ai, Li Wang, *Senior Member, IEEE*
Bingli Jiao, *Senior Member, IEEE*, and Kwang-Cheng Chen, *Fellow, IEEE*

***Abstract*—Recently, mobile cloud computing (MCC) is drawing substantial attention due to potential reduction on local energy consumption and local execution time. To further satisfy the requirement of latency-intensive applications in the future, fog computing is proposed to take full advantages of close-by nodes with available resources, thus decreasing the transmission time during computation offloading. This paper jointly considers near-by nodes and network edge base station in fog computing to provide a mutually complementary service, making the computation resources fully utilized. Furthermore, non-orthogonal multiple access (NOMA) has been introduced as a transmission technology to support high spectrum efficiency and low latency. Consequently, a new technology challenge regarding latency minimization can be facilitated as a computational offloading using NOMA. By further leveraging the concept of social networks, we therefore propose a social trust based NOMA cooperative computation offloading algorithm (SNOMA-COA) to minimize the completion time of system. Simulation results demonstrate that the proposed algorithm achieves the goal to effectively reduce system latency by offloading computation.**

***Index Terms*—Fog computing, non-orthogonal multiple access (NOMA), social trust.**

I. Introduction

With the popularity of mobile applications, it is feasible to resort the computation offloading technology to meet the requirement of the applications with highly complicated computation resources. By removing parts of computation-intensive tasks from local executing to the cloud with enough computing and storage resource, mobile cloud computing (MCC) computation offloading technology can reduce the execution time and energy consumption of mobile device, improving the performance of applications [1] [2]. Nevertheless, the long distance between the user and cloud will lead to excessive latency and consumption during the transmission process. Aiming at this problem, fog computing is put forward, utilizing the computing functions of close-by end users [3]. Specifically, when mobile devices (or fog nodes) deal with intensive computation beyond their own limited computation capacity and cannot resort the remote cloud due to the requirement of latency or consumption, offloading nodes can seek nearby available fog nodes to offload the intensive computation by cooperative communication in place of remote cloud, which can effectively support latency-sensitive functions.

Nowadays communications technology in the future fifth generation (5G) requires not only high data rate communications with ultra-low latency [4], but also higher spectrum efficiency because of spectrum shortage [5] [6]. In contrast to conventional orthogonal multiple access (OMA) where radio resource is exclusively allocated to each user, non-orthogonal multiple access (NOMA) allows allocating one frequency channel to multiple users within the same cell simultaneously and offers high spectrum efficiency [7] [8] [9]. Fully exploiting performance improved by NOMA on spectrum efficiency, [10] proposed a NOMA cluster strategy in downlink and uplink to enhance the sum-throughput of system. [11] combined NOMA with spatial modulation technology in wireless vehicle-to-vehicle (V2V) environments, greatly improving the system spectrum efficiency.

Traditional wireless networking traditionally treats computation offloading by assuming all cooperative nodes including those at the edge trustworthy. However, the interaction and thus social relationship of involved cooperative nodes are overlooked, particularly the offloading nodes and their willingness to assist other nodes at the edge of the wireless networks. Based on the interactive relationship, [12] exploited the social interaction influenced by social trust among network nodes to realize efficient and effective cooperative networking, where such social network information can improve system performance [13], since choosing reliable nodes can enhance communication link stability and probability of successful communication [13] [14] [15], meantime increasing the reliability of computation offloading.

Inspired by these factors, in this paper a scenario with multiple offloading nodes and cooperative nodes are considered in fog computation offloading. Unlike traditional computation offloading system, information in social domain besides physical domain is utilized to enhance system performance. NOMA technology is considered in transmission process to further boost system spectrum efficiency and decrease latency. Specifically, we formulate a latency minimization problem under transmission power, energy consumption, and NOMA clustering constraints. Furthermore, a social trust based NOMA cooperative computation offloading algorithm (SNOMA-COA) is proposed to solve this latency minimization problem.

L. Wang (*corresponding author*) and Y. Ai are with the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China (Email: {liwang, ytai}@bupt.edu.cn). L. Wang is also with National Mobile Communications Research Laboratory, Southeast University, Nanjing, China.

B. Jiao is with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China (Email: jiaobl@pku.edu.cn).

K. C. Chen is with the Department of Electrical Engineering, University of South Florida, Tampa, FL, USA (Email: kwangcheng@usf.edu).

The work is supported in part by the NSFC of China (Grants No. 61571056), and the open research fund of the National Mobile Communications Research Laboratory of Southeast University (Grant No. 2016D04), and the State Major Science and Technology Special Projects of China under Grant 2016ZX03001017-004.

978-1-5386-2062-5/17/$31.00 ©2017 IEEE

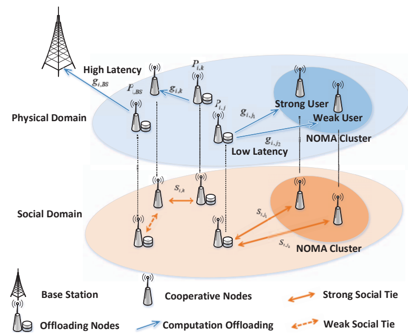


Fig. 1. The computation offloading scenario.

II. System Model And Problem Formulation

In this section, we consider a multi-user computation offloading system as shown in Fig. 1, where *offloading nodes* indicate the fog nodes with tasks needed to be offloaded. Fog nodes which help offloading nodes to compute partial tasks are defined as *cooperative nodes*. N fog nodes process the computation tasks beyond their own computation capacity, so partial tasks need to be offloaded to other P nodes. For simplicity, we only consider three ways to perform the computation offloading as shown in Fig. 1.

1) **Single cooperative node computation offloading:** single cooperative node help the offloading node i to compute partial computation tasks.
2) **NOMA cluster computation offloading:** one NOMA cluster is assigned to the offloading node i to compute partial computation tasks.
3) **Edge base station computation offloading:** edge base station performs partial computation tasks for offloading node i, which serves as a complementary computation offloading strategy if there is no qualified cooperative nodes to help offloading node i.

Assume that tasks of node i can be partitioned with one part executed locally and the other part offloaded to close-by trusted nodes or base station. Denote L_i as the input-data size (in bits) of tasks which need to be offloaded. X_i represents the computation workload/intensity (in CPU cycles per bits).

A. Single Cooperative Node Computation Offloading

In computation offloading scenario with the help of only one cooperative node, the received signal of cooperative node k from node i can be expressed as

$$y_k = \sqrt{P_{i,k}} x_{i,k} g_{i,k} + n_k, \tag{1}$$

and the data rate of the direct cooperative transmission link from node i to one cooperative node k can be expressed as

$$R_{i,k}^C = W \log_2 (1 + \frac{P_{i,k} g_{i,k}}{\sigma_k^2}), \tag{2}$$

where y_k is the received signal of cooperative node k. $x_{i,k}$, n_k and W are the transmitted signal, Gaussian noise, and channel bandwidth, respectively. $P_{i,k}$ is the transmission power for computation offloading from node i to node k. $g_{i,k}$ is the channel gain between the node i and the cooperative node k. σ_k^2 is the background Gaussian noise power. Since the size of computation results are much smaller than the size of input data generally, the time to send the results back to the offloading node could be neglected [16]. So the total task completion time for node i can be calculated as

$$T_{i,k}^C = T_{i,k}^t + T_{i,k}^e = \frac{L_i}{R_{i,k}^C} + \frac{L_i X_i}{f_{n,k}}, \tag{3}$$

where T_i^t, T_i^e are transmission time and execution time, respectively. $f_{n,k}$ represents the computational capability (i.e., CPU cycles per second) of cooperative node k. The energy consumption can be calculated as [17]

$$\begin{aligned} E_{i,k}^C &= E_{i,k}^t + E_{i,k}^e = P_{i,k} T_{i,k}^t + \kappa L_i X_i f_{n,k}^2 \\ &= P_{i,k} \frac{L_i}{R_i^C} + \kappa L_i X_i f_{n,k}^2, \end{aligned} \tag{4}$$

where κ is a constant related to the hardware architecture.

B. NOMA Cluster Computation Offloading

When the offloading node i can find no less than two trusted nodes, the channel condition will be estimated to ensure that they can form a NOMA cluster with two users. According to [7], the NOMA cluster can consist of a strong user with better channel condition and a weak user with relatively poor channel condition. To perform successive interference cancellation (SIC) successfully, transmission power for each NOMA user needs to be selected properly. Denote node j_1 and node j_2 as the strong user and weak user in NOMA cluster, respectively. Therefore, the transmission power allocated for the strong user is $\alpha P_{i,j}$, and the transmission power allocated for the weak user is $(1-\alpha)P_{i,j}$ where $\alpha \in [0, \frac{1}{2}]$ is a power allocation factor [11].

In NOMA, for strong user, SIC is performed first to decode the signal of weak user, and then the decoded signal of weak user is subtracted from the received signal. Eventually, the resultant signal is used to decode the signal for strong user. As for weak user, SIC is not executed, and the signal is directly decoded treating the signal of strong user as interference. Thus, the received signal of j_1 and j_2 are

$$y_{j_1} = \sqrt{\alpha P_{i,j}} x_{i,j_1} g_{i,j_1} + n_{j_1}, \tag{5}$$

$$y_{j_2} = \sqrt{\alpha P_{i,j}} x_{i,j_1} g_{i,j_2} + \sqrt{(1-\alpha) P_{i,j}} x_{i,j_2} g_{i,j_2} + n_{j_2}, \tag{6}$$

where x_{i,j_1} and x_{i,j_2} denote the transmitted signal for j_1 and j_2. n_{j_1} and n_{j_2} are Gaussian noise whose power are $\sigma_{j_1}^2$ and $\sigma_{j_2}^2$. $P_{i,j}$ is the transmission power of node i for computation offloading to the NOMA cluster j, g_{i,j_1} is the channel gain between the node i and the NOMA strong user j_1, and g_{i,j_2} is the channel gain between the node i and the NOMA weak user j_2.

The achievable data rate for strong user j_1 with higher signal to interference plus noise ratio (SINR) and weak user with lower SINR j_2 are given as

$$R_{i,j_1}^{NOMA} = W \log_2 (1 + \frac{\alpha P_{i,j} g_{i,j_1}}{\sigma_{j_1}^2}), \tag{7}$$

$$R_{i,j_2}^{NOMA} = W\log_2(1 + \frac{(1-\alpha)P_{i,j}g_{i,j_2}}{\alpha P_{i,j}g_{i,j_2} + \sigma_{j_2}^2}). \tag{8}$$

Hence the total data rate for the NOMA cluster j can be written as

$$R_{i,j}^{NOMA} = R_{i,j_1}^{NOMA} + R_{i,j_2}^{NOMA}. \tag{9}$$

Assume that $\beta_{i,j_1}L_i$ tasks are accomplished by the strong user j_1 and $\beta_{i,j_2}L_i = (1-\beta_{i,j_1})L_i$ tasks are accomplished by the weak user j_2 where $\beta_{i,j_1} \in [0,1]$ and $\beta_{i,j_2} \in [0,1]$ are input-data size allocation indices which are decided by channel conditions and computation capacities of cooperative nodes. The nodes with higher achievable data rate and computation capacity will be allocated more computation tasks. Since tasks for two nodes in a NOMA cluster are transmitted and computed simultaneously, so we assume that the tasks are completed until two nodes finish their computation tasks. Similar to the single cooperative node computation offloading, the total task completion time of NOMA cluster computation offloading can be expressed as

$$T_{i,j}^{NOMA} = \max\left(\frac{\beta_{i,j_1}L_i}{R_{i,j_1}^{NOMA}} + \frac{\beta_{i,j_1}L_iX_i}{f_{n,j_1}}, \frac{\beta_{i,j_2}L_i}{R_{i,j_2}^{NOMA}} + \frac{\beta_{i,j_2}L_iX_i}{f_{n,j_2}}\right), \tag{10}$$

where $\max(x,y)$ denote the maximum value of set. f_{n,j_1} and f_{n,j_2} are computation capability of node j_1 and j_2 in NOMA cluster, respectively.

The energy consumption of NOMA cluster computation offloading can be calculated as

$$\begin{aligned} E_{i,j}^{NOMA} &= \alpha P_{i,j}\frac{\beta_{i,j_1}L_i}{R_{i,j_1}^{NOMA}} + (1-\alpha)P_{i,j}\frac{\beta_{i,j_2}L_i}{R_{i,j_2}^{NOMA}} \\ &+ \kappa\beta_{i,j_1}L_iX_if_{n,j_1}^2 + \kappa\beta_{i,j_2}L_iX_if_{n,j_2}^2. \end{aligned} \tag{11}$$

C. Edge Base Station Computation Offloading

When offloading partial computation tasks to the edge base station, the received signal of base station from node i can be expressed as

$$y_{BS} = \sqrt{P_{i,BS}}x_{i,BS}g_{i,BS} + n_{BS}, \tag{12}$$

and the data rate of the transmission link between node i and base station can be expressed as

$$R_i^{BS} = W\log_2(1 + \frac{P_{i,BS}g_{i,BS}}{\sigma_{BS}^2}), \tag{13}$$

where y_{BS} is the received signal of base station and $x_{i,BS}$ is the transmitted signal. $P_{i,BS}$ is the transmission power of node i for computation offloading to the base station. $g_{i,BS}$ is the channel gain between the node i and base station. n_{BS} is Gaussian noise whose power is σ_{BS}^2.

The total task completion time of base station computation offloading can be expressed as

$$T_i^{BS} = \frac{L_i}{R_i^{BS}} + \frac{L_iX_i}{f_{BS}}. \tag{14}$$

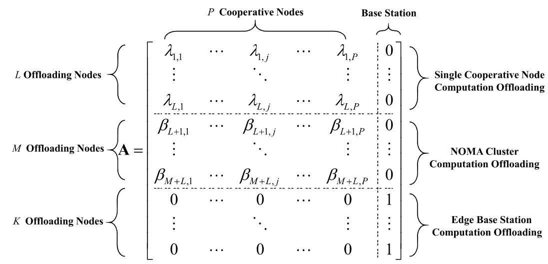


Fig. 2. Matrix representation of computational task offloading strategy.

The energy consumption of NOMA cluster computation offloading can be calculated as

$$E_i^{BS} = P_{i,BS}\frac{L_i}{R_i^{BS}} + \kappa L_iX_if_{BS}^2. \tag{15}$$

It should be noted that the computation capacity of base station is larger than the cooperative nodes. However, the latency of base station computation offloading is also higher than cooperative nodes computation offloading because of long distance.

Here, let matrix $\mathbf{A}_{N\times(P+1)}$ indicate the computation tasks offloading strategy where $\mathbf{A}$ is expressed as Fig. 2. Assume that L nodes in set $\mathcal{P}$ process single cooperative computation offloading and M nodes in set $\mathcal{Q}$ process NOMA cluster computation offloading. $K = N-L-M$ nodes in set $\mathcal{R}$ process base station computation offloading. L, M, and K are variable nonnegative integers. $\lambda_{i,j}$ is an input-data size allocation index in single node cooperative computation offloading, where $\lambda_{i,j} = 1$ indicates the i-th offloading node offload its whole computation tasks to the j-th cooperative node. $\beta_{i,j}$ denotes the input-data size allocation index which has been defined in (10). Again, it would be emphasized that when the i-th offloading node choose the j_1-th and j_2-th offloading node to perform computation offloading, $\beta_{i,j_1} + \beta_{i,j_2} = 1$.

As mentioned above, in order to select the optimal cooperative nodes, we take social characteristics of cooperative nodes into consideration, by introducing the social trust index $s_k \in [0,1]$, $s_{i,j_1} \in [0,1]$, and $s_{i,j_2} \in [0,1]$ to evaluate the reliable level of nodes. Nodes with higher social trust index indicate more reliable nodes. One cooperative nodes can communicate with at most one offloading node at the same time. Besides, we assume that during the process of computation offloading, the offloading node will not change its role until tasks are accomplished completely.

Thus the total data rate of the system can be expressed as

$$\begin{aligned} R_{tot} &= \sum_{i=1}^{L} s_k R_{i,k}^C + \sum_{i=1}^{L}(1-s_k)R_i^{BS} + \sum_{i=1}^{M}(s_{i,j_1}s_{i,j_2}R_{i,j}^{NOMA}) \\ &+ \sum_{i=1}^{M}((1-s_{i,j_2})s_{i,j_1}R_{i,j_1}^C + (1-s_{i,j_1})s_{i,j_2}R_{i,j_2}^C) \\ &+ \sum_{i=1}^{M}(1-s_{i,j_1})(1-s_{i,j_2})R_i^{BS} + \sum_{i=1}^{K}R_i^{BS}. \end{aligned} \tag{16}$$

It is worth mentioning that when any single one of cooperative nodes in NOMA cluster may not be reliable, and another node can still process single cooperative node computation offloading. When there is no trusted node, the offloading node will resort to the base station just as shown above.

Denote system latency as the total task completion time of all offloading nodes. Similarly, system latency and energy consumption of the system can be expressed as

$$\begin{aligned}T_{tot} = &\sum_{i=1}^{L} s_k T_{i,k}^{C} + \sum_{i=1}^{L}(1-s_k)T_i^{BS} + \sum_{i=1}^{M}(s_{i,j_1}s_{i,j_2}T_{i,j}^{NOMA}) \\ &+ \sum_{i=1}^{M}((1-s_{i,j_2})s_{i,j_1}T_{i,j_1}^{C} + (1-s_{i,j_1})s_{i,j_2}T_{i,j_2}^{C}) \\ &+ \sum_{i=1}^{M}(1-s_{i,j_1})(1-s_{i,j_2})T_i^{BS} + \sum_{i=1}^{K}T_i^{BS}, \end{aligned} \tag{17}$$

$$\begin{aligned}E_{tot} = &\sum_{i=1}^{L} s_k E_{i,k}^{C} + \sum_{i=1}^{L}(1-s_k)E_i^{BS} + \sum_{i=1}^{M}(s_{i,j_1}s_{i,j_2}E_{i,j}^{NOMA}) \\ &+ \sum_{i=1}^{M}((1-s_{i,j_2})s_{i,j_1}E_{i,j_1}^{C} + (1-s_{i,j_1})s_{i,j_2}E_{i,j_2}^{C}) \\ &+ \sum_{i=1}^{M}(1-s_{i,j_1})(1-s_{i,j_2})E_i^{BS} + \sum_{i=1}^{K}E_i^{BS}. \end{aligned} \tag{18}$$

In the following, an optimization problem $\mathbf{P}$ to minimize the total completion time of all N nodes in the cooperative communication system is formulated.

$$\textbf{P:} \min_{\mathbf{A}, L,\, M,\, K} T_{tot} \tag{19}$$

$$\text{s.t.} \max\{P_{i,k}, P_{i,j}, P_{i,BS}\} \leq P_{\max}, \tag{19a}$$

$$\min\left\{\frac{P_{i,k}g_{i,k}}{\sigma_k^2}, \frac{P_{i,j_1}g_{i,j_1}}{\sigma_{j_1}^2}, \frac{P_{i,j_2}g_{i,j_2}}{\sigma_{j_2}^2}, \frac{P_{i,BS}g_{i,BS}}{\sigma_{BS}^2}\right\} \geq \gamma_{\min}, \tag{19b}$$

$$\min\{s_{i,k}, s_{i,j_1}, s_{i,j_2}\} \geq s_{\min}, \tag{19c}$$

$$\min\{g_{i,k}, g_{i,j_1}, g_{i,j_2}\} \geq g_{\min}, \tag{19d}$$

$$g_{i,j_1} - g_{i,j_2} \geq g_{gap}, \tag{19e}$$

$$E_{tot} \leq E_{max}, \tag{19f}$$

$$0 \leq \beta_i \leq 1, \tag{19g}$$

$$0 \leq \alpha \leq \frac{1}{2}, \tag{19h}$$

$$L + M + K = N. \tag{19i}$$

Constraint (19a) ensures that the power allocated to the offloading node is less than maximum transmission power. Constraint (19b) ensures that signal-to-noise ratio (SNR) at the cooperative node should satisfy the minimum SNR requirement. Constraint (19c) and (19d) respectively ensure the requirement of channel gain and social trust, and g_{min}, s_{min} are the minimum requirement of channel gain and social trust between offloading nodes and cooperative nodes. Constraint (19e) is the requirement of forming a NOMA cluster where g_{gap} is the minimum channel gain difference. Constraint (19f) ensures that the total energy consumption is less than the total system energy.

III. Social Trust Based NOMA Cooperative Offloading Algorithm

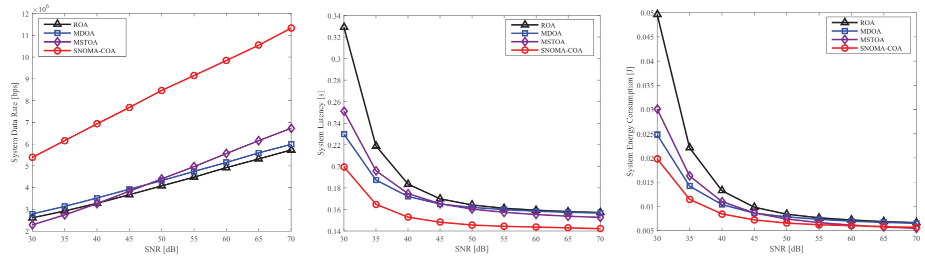
Algorithm 1: Social Trust Based NOMA Cooperative Offloading Algorithm (SNOMA-COA)

Input: input parameters N, P, g_{min}, s_{min}, g_{gap}

while $N \neq 0$ **do**
  For each offloading node i, find P' qualified cooperative nodes which can meet the requirement g_{min}, s_{min};
  if $P' \geq 2$ **then**
    if *There is two cooperative nodes satisfying NOMA cluster requirement (i.e. Constrain (19e))* **then**
      Offloading node i performs NOMA cluster computation offloading;
      Offloading node $i \in \mathcal{P}$, $P' = P' - 2$;
      Calculate the completion time in (10);
    else
      Offloading node i performs single cooperative node computation offloading;
      Offloading node $i \in \mathcal{Q}$, $P' = P' - 1$;
      Calculate the completion time in (3);
    end
  end
  if $P' = 1$ **then**
    Offloading node i performs single cooperative node computation offloading;
    Offloading node $i \in \mathcal{Q}$, $P' = P' - 1$;
    Calculate the completion time in (3);
  end
  if $P' = 0$ **then**
    Offloading node i performs base edge station computation offloading;
    Offloading node $i \in \mathcal{R}$, $P' = P'$;
    Calculate the completion time in (14);
  end
  Delete the offloading node with least completion time and its cooperative nodes;
  $N = N - 1$;
end

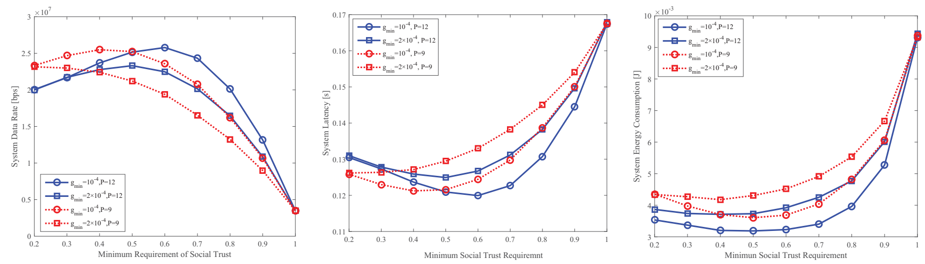
In this section, social trust will be introduced to NOMA cooperative offloading algorithm to minimize the latency. In SNOMA-COA, N and P are respectively the number of offloading nodes and cooperative nodes. Since NOMA can greatly improve the spectrum efficiency of system, so node i prefers to offload its computation tasks to NOMA cluster instead of single cooperative node. Besides, node i prefers to cooperate with nodes around it rather than offload computation tasks to the base station since base station is far away from offloading nodes so that the latency is long.

Specifically, as shown in the **algorithm 1**, the node that cannot satisfy the requirement of social trust and channel gain will be deleted first. In most cases, social trust requirement in SNOMA-COA can guarantee that offloading nodes can choose



(a) System data rate (b) System latency (c) System energy consumption

Fig. 3. System data rate, system latency, system energy consumption of different offloading algorithms.



(a) System data rate (b) System latency (c) System energy consumption

Fig. 4. System data rate, system latency, system energy consumption of different minimum requirement of social trust s_{min} with $g_{min} = 10^{-4}, 2 \times 10^{-4}$ and the number of cooperative nodes $P = 9, 12$ in SNOMA-COA.

reliable cooperative nodes and hence the success possibility of computation offloading can be enhanced.

Motivated by the analysis in section II, when there is no trusted node around the node i, node i will offload its computation task to the edge base station with long latency. If there is only one trusted node around node i, node i will offload its partial computation tasks to it (i.e. single cooperative offloading mode). When there are at least two trusted nodes around node i, they will be checked whether they can form a NOMA cluster made up of two nodes. In other words, the channel gain of selected nodes in NOMA cluster should not be similar so that their own information can be decoded successfully. Otherwise, one node which can improve the data rate of the system most will be selected to help the node i to process the computation offloading. To emphasize our proposed SNOMA-COA in **Algorithm 1**, other three algorithms are considered as benchmarks as follows.

Random Offloading Algorithm (ROA): Without offloading qualification judgement (i.e. Constrain (19c) and (19d)), offloading nodes choose their cooperative nodes randomly.

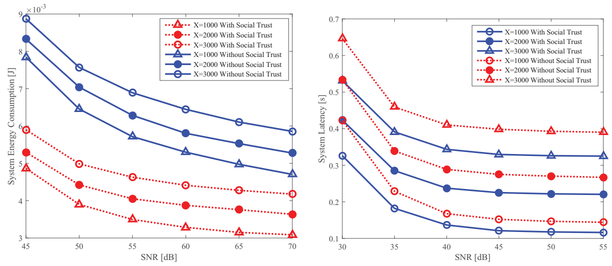
Minimum Distance Offloading Algorithm (MDOA): Offloading nodes will choose the nearest cooperative nodes to perform the computation offloading in physical domain.

Maximum Social Trust Offloading Algorithm (MSTOA): Offloading nodes will choose the most reliable cooperative nodes to perform the computation offloading in social domain with the requirement g_{min} guaranteed.

IV. Simulation Results

Consider a cooperative offloading system where there are $N = 5$ offloading nodes and $P = 11$ cooperative nodes, and all nodes are randomly deployed in a 1km-radius cell with single antenna. Base station is located 10km away from the cell. Large scale pathloss is considered where $PL = d^{-\theta}(\theta = 2)$ and d is the distance between nodes. The channel bandwidth W is 1MHz and input-data size L is 10^5 bits. NOMA power allocation factor α is 0.3, and g_{gap} is 2×10^{-4}. Computation workload X is 1000 cycles/bits. The transmission power and noise power is 250mW and -96dBm, respectively. CPU computing capacity of nodes f_n and base station f_{BS} are 700MHz and 755MHz [17]. For equal comparison between offloading strategies with NOMA and without NOMA, assume that every cooperative node uses $\frac{1}{2}W$ bandwidth to communicate with offloading nodes [7].

Fig. 3 shows the comparison with other three offloading algorithms. As these figures show, ROA performs worst because offloading nodes are likely to offload their tasks to unreliable nodes or nodes with poor channel conditions. MDOA performs better than ROA because this algorithm can guarantee the channel conditions between offloading nodes and cooperative nodes are better than ROA in most cases. It is noted that the MSTOA performs better than ROA and MDOA and consumes almost same energy as our proposed algorithm with high SNR. Because when SNR is high, the latency and energy consumption of transmission process is much less than that in tasks execution process so social trust index will be the dominant factor. Our proposed SNOMA-COA achieves a much better performance of system data rate than other three algorithms since NOMA can greatly improve the spectrum efficiency. Besides, system latency will also be lower because tasks can be computed by two cooperative nodes in NOMA cluster simultaneously.



(a) System energy consumption (b) System latency

Fig. 5. System energy consumption and system latency of different SNR with computation workload $X = 1000$, 2000, 3000 in SNOMA-COA.

Fig. 4 presents the system data rate, system latency, and system energy consumption with different s_{min}, g_{min} and different numbers cooperative nodes P in SNOMA-COA. It can be seen that the system performance is improved with the increase of s_{min} at first because s_{min} can guarantee that some unreliable cooperative nodes are not considered to perform the computation offloading. However, higher s_{min} is not necessarily related to better performance. With high s_{min}, more offloading nodes will not find cooperative nodes so that they have to offload their tasks to edge base station, resulting in lower data rate, longer latency, and higher energy consumption.

Besides, the system performance decreases with the increase of g_{min} because less cooperative nodes can be chosen to perform the computation offloading so that offloading nodes have to offload their tasks to base station. However, if g_{min} is too small, the quality of communications will not be guaranteed due to extraordinarily poor channel condition.

It also can be noticed that system performance with less cooperative nodes is better at the low social trust requirement because more cooperative nodes will increase the possibility of offloading to unreliable nodes. However, as s_{min} increases, unreliable nodes are not chosen to help the offloading nodes, and hence system performance can be improved since more reliable cooperative nodes can be chosen to help offloading.

Fig. 5 present system energy consumption and system latency with different SNR and different computation workloads in SNOMA-COA. As SNR increases, system data rate will increase so the transmission time decrease, which also results in less energy consumption in transmission process. Gradually, the system latency and energy consumption remain unchanged because tasks execution process also consume time and energy which are not affected by SNR. Besides, Higher computation workload X places more burdens on execution process, leading to longer system latency and higher system energy consumption. It should be noticed that the system performance can be further improved after considering the social trust.

V. Conclusion

In this paper, close-by nodes and base station at the network edge have been jointly resorted in fog computation offloading system. Based on social trust, a NOMA cooperative computation offloading algorithm called SNOMA-COA has been proposed to solve the latency minimization problem. Exploiting NOMA technology in cooperative communication has increased the system data rate and reduced the system latency. Moreover, after introducing social trust in computation offloading system, the link stability has been further enhanced. By utilizing the interplay between social characteristics, computation resources, and power resources into NOMA cooperative offloading system, improved system performance has been achieved in terms of stronger robustness, better spectrum efficiency, and lower latency.

References

[1] J. A. Suradkar and R. D. Bharati, "An effective computation offloading in pervasive devices to cloud," in *Proc. 2016 International Conference on Computing, Analytics and Security Trends (CAST)*, Pune, India, 2016, pp. 216-221.

[2] Liu, X. Zeng, W. Huang, J. Lin, X. Chen, and W. Guo, "Framework for context-aware computation offloading in mobile cloud computing," in *Proc. 2016 15th International Symposium on Parallel and Distributed Computing (ISPDC)*, Fuzhou, China, 2016, pp. 172-177.

[3] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854-864, Dec. 2016.

[4] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458-461, Mar. 2016.

[5] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36-43, May. 2014.

[6] S. Chen, F. Qin, B. Hu, X. Li and Z. Chen, "User-centric ultra-dense networks for 5G: challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 78-85, Apr. 2016.

[7] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys and Tut.*, vol. 19, no. 2, pp. 721-742, Secondquarter 2017.

[8] Z. Ding, Z. Zhao, M. Peng, and H. V. Poor, "On the spectral efficiency and security enhancements of NOMA assisted multicast-unicast streaming," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3151-3163, Jul. 2017.

[9] Y. Chen, L. Wang, and B. Jiao, "Cooperative multicast non-orthogonal multiple access in cognitive radio," in *Proc. 2017 IEEE International Conference on Communications (ICC)*, Paris, France, Jul. 2017, pp. 1-6.

[10] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, no. 1 , pp. 6325-6343, Aug. 2016.

[11] Y. Chen, L. Wang, Y. Ai, B. Jiao, and L. Hanzo, "Performance analysis of NOMA-SM in vehicle-to-vehicle massive MIMO channels," *IEEE J. Sel. Areas in Commun. (JSAC)*, vol. PP, no. 99, pp.1-13, Jul. 2017.

[12] X. Chen, B. Proulx, X. Gong, and J. Zhang, "Exploiting social ties for cooperative D2D communications: A mobile social networking case, *IEEE/ACM Trans. Netw.*, vol. 23, no. 5, pp. 1471-1484, Oct. 2015.

[13] L. Wang, H. Tang, and M. Cierny, "Device-to-device link admission policy based on social interaction information," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4180-4186, Sep. 2015.

[14] L. Wang, H. Wu, and G. L. Stüber, "Cooperative jamming-aided secrecy enhancement in P2P communications with social interaction constraints," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1144-1158, Feb. 2017.

[15] L. Wang, H. Wu, W. Wang, and K. C. Chen, "Socially enabled wireless networks: resource allocation via bipartite graph matching," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 128-135, Oct. 2015.

[16] D. Huang, P. Wang, and D. Niyato, "A dynamic offloading algorithm for mobile computing," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 1991-1995, Jun. 2012.

[17] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, Mobile edge computing: Survey and research outlook, *submitted to IEEE Commun. Surveys Tuts.*, Jan. 2017. [Online]. Available: https://arxiv.org/pdf/1701.01090v1.pdf